

Feedback Solution to Optimal Switching Problems with Switching Cost

Ali Heydari¹

Abstract—The problem of optimal switching between nonlinear autonomous subsystems is investigated in this study where the objective is not only bringing the states to close to the desired point, but also adjusting the switching pattern, in the sense of penalizing switching occurrences and assigning different preferences to utilization of different modes. The mode sequence is unspecified and a switching cost term is used in the cost function for penalizing each switching. It is shown that once a switching cost is incorporated, the optimal cost-to-go function depends on the already active subsystem, i.e., the subsystem which was engaged in the previous time step. Afterwards, an approximate dynamic programming based method is developed which provides an approximation of the optimal solution to the problem in a *feedback* form and for *different initial conditions*. Finally, the performance of the method is analyzed through numerical examples.

I. INTRODUCTION

Many real-world control problems can be classified as switching problems in the sense that the system subject to control is comprised of several different modes (sometimes called subsystems) and at each instant only one of the modes can be active. A basic example of such a system is a plant equipped with on-off actuators [1]. The solution to such problems includes a *switching schedule* which determines the number of switching, the switching instants, and the order of the active subsystems.

The developments in the field of optimal switching can be divided into different categories, two of which are nonlinear programming based methods and discretization based methods. Nonlinear programming based methods utilize the gradient of the cost with respect to the switching instants to calculate local optimal switching times using nonlinear programming [2]–[9]. In these methods, the sequence of active subsystems, known as the *mode sequence*, is typically selected a priori. The problem is then simplified to determining the switching instants between the modes. Discretization based methods, however, discretize the state and input space to end up with a finite number of choices [10], [11]. Among the intelligent approaches to the problem, genetic algorithm and neural networks were used in Refs. [12] and [13], respectively, to determine the optimal switching for one set of initial conditions.

All the cited methods work only with a specific initial condition; each time the initial condition is changed, a new set of computations needs to be performed to find the new optimal switching instants. In order to extend the validity of the results for different initial conditions within a pre-selected set, in [5]

a solution was found as the local optimum in the sense that it minimizes the worst possible cost for all trajectories starting in the selected initial states set. Another drawback of majority of the methods, especially the nonlinear programming based methods, is the fact that they lead to an open loop solution.

On the other hand, approximate dynamic programming (ADP) has shown great potentials in solving conventional optimal control problems with infinite-horizon cost functions [14]–[20] and also with finite-horizon cost functions [21]–[23]. The backbone of the ADP based methods is using Bellman equation [24] and approximating the mapping between the states of the system and the optimal control. These potentials motivated the author of this study to investigate the application of ADP to *switching* problems in his PhD research. This was done through developing solutions to problems with *fixed mode sequence* and *fixed number of switching* [25], [26], problems with *free mode sequence* and *controlled subsystems* [27], and also problems with *free mode sequence* and *autonomous subsystems* [28]. The interesting feature of these developments is the fact that they provide approximate optimal solution for a vast domain of initial conditions. Another advantage of these methods is their feedback nature. These developments, however, do not provide the designer with the ability of influencing, e.g., decreasing the number of switching. For example, in the extreme case, the solutions proposed in [27] and [28] can lead to one switching at every single sampling time. Such a switching frequency is impracticable in many applications. Different tricks, however, are proposed in [27] and [28] for manipulating the switching frequency. These remedies lead to deviation of the solution from optimality and can potentially destabilize the system. Moreover, these developments assume a cost function which is independent of the active mode, hence, the designer cannot assign different costs (or preferences) to different modes.

In an independent study in utilizing ADP for solving optimal switching problems, the authors of [29] proposed a method for solving switching problems with finite-horizon cost functions. The proposed method, however, inherits the curse of dimensionality from dynamic programming, in the sense that, at each iteration of the learning process as many cost-to-go functions as the number of subsystems raised to the power of the iteration number should be learned. For example, for a three mode system, at 100th iteration, the number of functions subject to learning is 3^{100} . Moreover, the proposed training algorithm is based on a selected initial state.

Another investigation for solving switching problems using ADP was recently reported in [30] with a different approach. However, initial conditions are assumed to be known a priori and the result does not admit penalizing each switching. These two points differentiate the work from this study.

¹Assistant Professor of Mechanical Engineering, South Dakota School of Mines and Technology, Rapid City, SD, email: ali.heydari@sdsmt.edu.

An idea for influencing (decreasing) the number of switching is incorporating a *switching cost* term in the cost function, for the purpose of penalizing each switching between the modes, [11]. Moreover, utilizing a cost function with mode dependent terms, i.e., having a *switching cost function*, leads to the desired feature of assigning different costs to different modes. These modifications, however, lead to a very important change in the characteristics and the nature of the solution. It is shown in this study that in case of penalizing each switching, the optimal cost-to-go becomes a function of the subsystem which was active at the *previous* time step, i.e., the *already* active subsystem. Consequently, the methods reported in [25]-[28] fail to provide solutions to problems with such cost functions. Based on the developments in [25]-[28], this study is aimed at developing a new switching method which admits the switching cost term and the switching cost function. This is the main contribution of this paper and is carried out through a new switching law, a new neural network (NN) structure as the function approximator, and a new parameter/weight update algorithm. Afterwards, the continuity of the function subject to approximation is analyzed and certain changes in the selected NN form is proposed for satisfying the necessary condition for uniform approximation of the desired function. Finally, the performance of the method is analyzed numerically in different examples.

The closest study in the literature to the problem subject to this paper is [11]. The differences are a) a maximum number of switching needs to be assumed, b) the state space needs to be discretized, and c) the solution needs to be calculated numerically in [11]. In this study, however, the number of switching is free, to be obtained such that the cost function is minimized, the state vector can change continuously, and the (approximate) solution is calculated in a closed form.

The rest of this paper is organized as follows. The problem is formally presented in section II and the proposed solution is detailed in section III. Section IV discusses the online implementation of the proposed method and section V includes the numerical analyses and simulations. Finally, the conclusions are given in section VI.

II. PROBLEM FORMULATION

The dynamics of the M autonomous modes/subsystems of a switching system can be modeled using

$$x_{k+1} = f_i(x_k), k \in \mathcal{K}, i \in \mathcal{I} \quad (1)$$

where $f_i : \mathbb{R}^n \rightarrow \mathbb{R}^n, \forall i \in \mathcal{I} := \{1, 2, \dots, M\}$, $\mathcal{K} := \{0, 1, \dots, N-1\}$, and positive integer n is the dimension of the state vector x_k . Sub-index k in x_k represents the discrete time index and sub-index i in $f_i(\cdot)$ represents the respective mode/subsystem. Denoting the active mode at instant k with i_k , a *switching schedule* identifies $i_k, \forall k \in \mathcal{K}$. Once a switching schedule is selected, the system can operate from the initial time $k = 0$ to the fixed final time $k = N$. The problem is defined as finding a switching schedule that minimizes the cost function given by

$$J = \psi(x_N, i_{N-1}) + \sum_{k=0}^{N-1} (Q(x_k, i_k) + \kappa(i_{k-1}, i_k)) \quad (2)$$

Cost function (2) is composed of three type of terms. a) Piecewise convex function $\psi : \mathbb{R}^n \times \mathcal{I} \rightarrow \mathbb{R}$ penalizes the error between the desired state value and the actual state value at the final time and is dependent on the active mode or configuration with which the operation of the system finishes, i.e., i_{N-1} . b) Piecewise convex function $Q : \mathbb{R}^n \times \mathcal{I} \rightarrow \mathbb{R}$ assigns different costs to the state error (the difference between the actual value and the desired value for the state vector) during the horizon and is dependent on the active mode during the horizon. c) Piecewise constant function $\kappa : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$ represents the switching cost. Each switching from mode i_{k-1} to i_k at time k leads to the cost represented by $\kappa(i_{k-1}, i_k)$, [11]. Therefore, $\kappa(i, i) = 0, \forall i \in \mathcal{I}$. For notational consistency in (2), the already active mode before the start of the process, i.e., before $k = 0$, is denoted with i_{-1} .

Remark 1: Functions $\psi(\cdot, i)$ and $Q(\cdot, i)$ are assumed to be convex, $\forall i \in \mathcal{I}$. Moreover, no assumption on the signs of the outputs of $\psi(\cdot, \cdot)$, $Q(\cdot, \cdot)$, and $\kappa(\cdot, \cdot)$, e.g., being positive semi-definite, is made and the theory developed in this study admits negative costs, i.e., rewards, as well.

III. PROPOSED SOLUTION

The method proposed in this study for solving the problem is based on approximating the *optimal cost-to-go*, i.e., the total cost from the current time to the final time, assuming the optimal decisions are made in selecting the modes for operating the system during the horizon. The optimal cost-to-go is sometimes called *value function* by some researchers. It is straightforward to see that the optimal cost-to-go is a function of the current state, i.e., x_k . Since the final time is fixed, the cost-to-go will depend on the current time as well. In other words, having the same current state, but a different *time-to-go*, i.e., different $N-k$, may lead to a different cost-to-go [24], [21]. Note that, the dependency on $N-k$ is equivalent of dependency on k , because, N is fixed and known.

An important observation for developing a solution to the problem defined in section II is the fact that the optimal cost-to-go also depends on the *previous active subsystem*, i.e., the subsystem which was active at the previous time, in problems with switching costs. The previous active subsystem, i_{k-1} , is ‘already’ active, hence, utilizing it at the current time step does not cause a switching cost, because $\kappa(i_{k-1}, i_{k-1}) = 0$. To see this dependency one may consider the difference between the following two example scenarios in controlling a switching system: a) the previous active subsystem is the same as the subsystem that the controller wants to activate at the current time, and b) having the same time k and current state x_k as in case ‘a’, the previous active subsystem is different than the one the controller wants to activate at the current time. Comparing these two scenarios it is seen that the optimal cost-to-go will be different due to the required switching in scenario ‘b’ and the respective incurred switching cost. Hence, the solution and the optimal cost-to-go at each instant are dependent on the previous active subsystem, as well as on the current time and state. Considering these dependencies, one may denote the optimal cost-to-go with $V_k^*(x_k, i_{k-1})$. Note that the sub-index k in $V_k^* : \mathbb{R}^n \times \mathcal{I} \rightarrow \mathbb{R}$ corresponds to the time dependency of the optimal cost-to-go.

Considering this concept, it is seen that the initial condition on the active mode, that is the active mode/configuration right before the start of the operation, plays a role in the selection of i_0 , when a switching cost is incorporated. In other words, depending on what the already active mode/configuration before the start of the process is, i.e., i_{-1} , the system may select a different i_0 . Therefore, as expected, the summation included in cost function (2) contains i_{-1} .

Remark 2: Another way of looking at the dependency of the cost-to-go on the already active subsystem, is considering the active subsystem/mode as a *state* of the system. In this case, a new state vector $X_k := [x_k^T, i_{k-1}]^T$ may be defined to represent the *overall* state of the system. This approach is also compatible with the physical way of looking at the modes as different *configurations* of the system. The active configuration, e.g., the position of the gear stick in a manual transmission car, is a physical state of the system.

Remark 3: The ‘already’ active mode should be differentiated from the ‘current’ active mode at time k . The former is the mode which was utilized at the ‘previous’ time step and is denoted with i_{k-1} , but, the latter is the mode which is going to be selected at the ‘current’ time step to operate the system from k to $k+1$ and is denoted with i_k . Following this terminology, the cost-to-go depends on the ‘previous’ (or the already active) subsystem, not on the ‘current’ subsystem.

A. Theory

The selected cost function, Eq. (2), leads to

$$V_N^*(x_N, i_{N-1}) = \psi(x_N, i_{N-1}), \forall i_{N-1} \in \mathcal{I}, \quad (3)$$

and

$$\begin{aligned} V_k^*(x_k, i_{k-1}) = & \psi(x_k^*, i_{N-1}^*) + \\ & (Q(x_k, i_k^*) + \kappa(i_{k-1}, i_k^*)) + \\ & \sum_{j=k+1}^{N-1} (Q(x_j^*, i_j^*) + \kappa(i_{j-1}^*, i_j^*)), \forall k \in \mathcal{K}, \forall i_{k-1} \in \mathcal{I}. \end{aligned} \quad (4)$$

where the *optimal* active mode at each instant j is denoted with i_j^* , $\forall j \in \mathcal{K}$, and the resulting optimal future states, calculated from (1), are denoted with x_j^* , $\forall j \in \mathcal{K} \cup \{N\}$. Eq. (4) can be formed as a recursive equation as

$$\begin{aligned} V_k^*(x_k, i_{k-1}) = & Q(x_k, i_k^*) + \kappa(i_{k-1}, i_k^*) + \\ & V_{k+1}^*(x_{k+1}^*, i_k^*), \forall k \in \mathcal{K}, \forall i_{k-1} \in \mathcal{I}, \end{aligned} \quad (5)$$

where $x_{k+1}^* = f_{i_k^*}(x_k)$. By the Bellman principle of optimality [24], one has

$$\begin{aligned} V_k^*(x_k, i_{k-1}) = & \min_{i \in \mathcal{I}} \left(Q(x_k, i) + \kappa(i_{k-1}, i) + \right. \\ & \left. V_{k+1}^*(f_i(x_k), i) \right), \forall k \in \mathcal{K}, \forall i_{k-1} \in \mathcal{I}. \end{aligned} \quad (6)$$

Moreover, the optimal mode i_k^* , which is also a function of k , x_k , and i_{k-1} , is given by

$$\begin{aligned} i_k^*(x_k, i_{k-1}) = & \operatorname{argmin}_{i \in \mathcal{I}} \left(Q(x_k, i) + \kappa(i_{k-1}, i) + \right. \\ & \left. V_{k+1}^*(f_i(x_k), i) \right), \forall k \in \mathcal{K}, \forall i_{k-1} \in \mathcal{I}. \end{aligned} \quad (7)$$

In other words, i_k^* is selected considering the following concerns:

- Selecting i_k^* leads to incurring the running cost of $Q(x_k, i_k^*)$.
- Selecting i_k^* leads to the next state vector being $f_{i_k^*}(x_k)$.
- Selecting i_k^* leads to the fact that at the next step, the already active subsystem will be i_k^* .
- Selecting i_k^* may lead to some switching cost due to i_k^* not being the same as the already active subsystem, which is i_{k-1} .

The first concern is addressed through the inclusion of $Q(x_k, i)$ in the minimization of Eq. (7). The second and third concerns are addressed through minimizing $V_{k+1}^*(f_i(x_k), i)$ in the right hand side of Eq. (7) with respect to its both i s. The fourth concern, however, is addressed through the term $\kappa(i_{k-1}, i)$ subject to minimization in (7). The existence of this term in Eq. (7) confirms the fact that the selection of each mode depends on the active mode at the previous step, as expected.

The key to the solution of the problem is the fact that if the optimal cost-to-go function $V_k^*(\cdot, \cdot)$ is calculated in a closed form for all $k \in \mathcal{K}$ then one can find the optimal i_k^* in a *feedback* form in online operation, as seen in (7). Motivated by the development in the ADP literature for conventional [14]-[23] and switching [25]-[28] problems, it is proposed to use a neural network (NN) as a function approximator for approximating the optimal cost-to-go function. Selecting a linear-in-parameter NN, the function is approximated within a compact set $\Omega \subset \mathbb{R}^n$ (representing the domain of interest) using

$$\begin{aligned} W_k^T \phi(x_k, i_{k-1}) \approx & V_k^*(x_k, i_{k-1}), \\ \forall k \in \mathcal{K} \cup \{N\}, \forall i_{k-1} \in \mathcal{I}, \forall x_k \in \Omega, \end{aligned} \quad (8)$$

where the selected smooth basis functions are given by $\phi : \mathbb{R}^n \times \mathcal{I} \rightarrow \mathbb{R}^m$, with m being a positive integer denoting the number of neurons. Unknown matrix $W_k \in \mathbb{R}^m$, to be found using learning algorithms, is the *weight* matrix of the network at time step k . Note that the time-dependency of the optimal cost-to-go function is incorporated using a NN with time-dependent weights. Moreover, the inputs to the basis functions correspond to the other dependencies of the function subject to approximation.

Before proceeding to the training algorithm, there is a concern with the selected NN structure (8) that needs to be resolved. It should be noted that NNs with continuous neurons are suitable for approximation of continuous functions [31], [32]. Otherwise, the approximation is not guaranteed to be uniform. Looking at (5), function $V_k^*(x_k, i_{k-1})$ may not be a continuous function versus i_{k-1} . As a matter of fact, since i_{k-1} belongs to a set of discrete integers, i.e., \mathcal{I} , it will not change continuously, therefore, the cost-to-go function also does not continuously change versus i_{k-1} , unless the system is comprised of only one mode. Hence, the selected network structure given in (8), with continuous basis functions $\phi(\cdot, \cdot)$ is not desired and a new structure should be used for implementation of the proposed solution. To remedy the problem, an innovative idea is proposed here, that is, using

NNs with i_{k-1} dependent weights for incorporation of i_{k-1} -dependency of the function subject to approximation. Let a new NN structure given by

$$W_k^{i_{k-1}^T} \phi(x_k) \approx V_k^*(x_k, i_{k-1}), \quad (9)$$

$$\forall k \in \mathcal{K} \cup \{N\}, \forall i_{k-1} \in \mathcal{I}, \forall x_k \in \Omega,$$

be used, where $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the selected set of basis functions and $W_k^i \in \mathbb{R}^m, \forall k \in \mathcal{K} \cup \{N\}, \forall i \in \mathcal{I}$, is the unknown weight matrix. Using the form given by (9), the number of weights required to be trained will be multiplied by the number of subsystems, as compared to the case of using the NN form given by (8).

Following this idea, it needs to be proved that each $V_k^*(x_k, i_{k-1})$ for every given i_{k-1} and k is a continuous function of x_k . Since, each one of such functions is being approximated using a different set of weights, denoted with $W_k^{i_{k-1}}$, it can be looked at as if each function $V_k^*(\cdot, i_{k-1})$ for every given k and i_{k-1} is being approximated using a separate NN denoted with $W_k^{i_{k-1}^T} \phi(\cdot)$. Therefore, the proof of continuity versus x_k suffices for having the desired uniform approximation capability for the NN structure given in (9). Theorem 1 proves the required continuity. While the main idea of the proof is adapted from [28], many changes are carried out to adapt the result for the cost-to-go function subject to the current study and to make the proof more rigorous.

Theorem 1: The optimal cost-to-go or value function for the problem of minimizing cost function (2) with respect to dynamics (1) is a continuous function of states in every compact set Ω , if functions $f_i(\cdot), \psi(\cdot, i)$, and $Q(\cdot, i), \forall i \in \mathcal{I}$ are continuous.

Proof: Function $V_N^*(\cdot, i)$ is continuous by Eq. (3) and the continuity of $\psi(\cdot, i), \forall i$. Assuming continuity of $V_{k+1}^*(\cdot, i), \forall i$, if it can be shown that function $V_k^*(\cdot, i)$ for all i s will be continuous, the proof is complete, by mathematical induction.

Let the scalar function $F : \mathbb{R}^n \times \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$ be defined as

$$F(x, j, i) := Q(x, i) + \kappa(j, i) + V_{k+1}^*(f_i(x), i), \quad (10)$$

and the piecewise constant function $i_k^* : \mathbb{R}^n \times \mathcal{I} \rightarrow \mathcal{I}$ be given by

$$i_k^*(x, j) = \operatorname{argmin}_{i \in \mathcal{I}} (F(x, j, i)). \quad (11)$$

It can be seen that function $F(x, j, i_k^*(x, j))$ is identical to $V_k^*(x, j)$ considering (5), (10), and (11). Therefore, continuity of $F(\cdot, j, i_k^*(\cdot, j))$ for all j s completes the proof.

Let \bar{x} be any selected point in Ω , for any given $j \in \mathcal{I}$ set

$$\bar{i} = i_k^*(\bar{x}, j). \quad (12)$$

Select an open set $\alpha \subset \Omega$ such that \bar{x} belongs to the boundary of α and limit

$$\hat{i} = \lim_{\|x - \bar{x}\| \rightarrow 0, x \in \alpha} i_k^*(x, j), \quad (13)$$

exists, where $\|\cdot\|$ denotes a vector norm. If $\bar{i} = \hat{i}$, for every such α , then there exists some open set $\beta \subset \Omega$ containing \bar{x} such that $i_k^*(x, j)$ is constant for all $x \in \beta$, because $i_k^*(x, j)$ only assumes integer values. In this case the continuity of $F(x, j, i_k^*(x, j))$ at $x = \bar{x}$ follows from the fact that $F(x, j, i)$

is continuous at $x = \bar{x}$, for every fixed $i \in \mathcal{I}$, by composition. The reason is $Q(\cdot, i)$, $f_i(\cdot)$, and $V_{k+1}^*(\cdot, i)$ are continuous functions and $\kappa(j, i)$ is a constant. Finally, the continuity of the function subject to investigation at every $\bar{x} \in \Omega$, leads to the continuity of the function in Ω .

Now assume $\bar{i} \neq \hat{i}$, for some α . From the continuity of $F(x, j, \hat{i})$ for the given \hat{i} , one has

$$F(\bar{x}, j, \hat{i}) = \lim_{\delta x \rightarrow 0} (F(\bar{x} + \delta x, j, \hat{i})) \quad (14)$$

If it can be shown that, for every selected α , one has

$$F(\bar{x}, j, \bar{i}) = F(\bar{x}, j, \hat{i}), \quad (15)$$

then the continuity of $F(x, j, i_k^*(x, j))$ versus x follows, because from (14) and (15) one has

$$F(\bar{x}, j, \bar{i}) = \lim_{\delta x \rightarrow 0} (F(\bar{x} + \delta x, j, \hat{i})), \quad (16)$$

and (16) leads to the continuity by definition [33]. The proof that (15) holds is done by contradiction. Assume that for some \bar{x} and some α one has

$$F(\bar{x}, j, \bar{i}) < F(\bar{x}, j, \hat{i}), \quad (17)$$

then, due to the continuity of both sides of (17) at \bar{x} for the fixed \bar{i} and \hat{i} , there exists an open set γ containing \bar{x} , such that

$$F(x, j, \bar{i}) < F(x, j, \hat{i}), \forall x \in \gamma. \quad (18)$$

Inequality (18) implies that at points which are *close enough* to \bar{x} , one has $i_k^*(x, j) \neq \bar{i}$. But, this contradicts Eq. (13) which implies that there always exists a point x *arbitrarily close* to \bar{x} at which $i_k^*(x, j) = \bar{i}$. Therefore, equality (18) cannot hold. Now, assume that

$$F(\bar{x}, j, \bar{i}) > F(\bar{x}, j, \hat{i}), \quad (19)$$

Inequality (19) leads to $i_k^*(\bar{x}, j) \neq \bar{i}$. But, this is against (12), hence, (19) also cannot hold. Therefore, (15) holds and hence, $F(x, j, i_k^*(x, j))$ is continuous at every $x \in \Omega$ for every fixed $j \in \mathcal{I}$. This completes the proof.

B. Training Algorithms

Selecting the NN structure, the next step is developing an algorithm for finding the unknown weights. Using Eqs. (3) and (6) the training algorithm can be derived in a *backward* fashion. Considering (3), unknown W_N can be obtained, for example using least squares method, as shown in [21]. Once W_N is found, Eq. (6) can be used for calculating W_{N-1} . Repeating this process, all the weights can be found from $k = N$ to $k = 0$. The training can be done either in a *batch* form or in a *sequential* form. Algorithm 1 details the batch training and Algorithm 2 presents the training in the sequential form.

Algorithm 1 - Batch Training

- Step 1: Randomly select p different state vectors $x^{[q]} \in \Omega, q \in \{1, 2, \dots, p\}$, for p being a large positive integer, where $\Omega \subset \mathbb{R}^n$ represents the domain of interest.
- Step 2: For $j = 1$ to M repeat Step 3.

Step 3: Find W_N^j such that

$$W_N^{jT} \phi(x^{[q]}) \approx \psi(x^{[q]}, j), \forall q \in \{1, 2, \dots, p\}. \quad (20)$$

Step 4: Set $k = N - 1$.

Step 5: For $j = 1$ to M repeat Step 6.

Step 6: Find W_k^j such that

$$W_k^{jT} \phi(x^{[q]}) \approx \min_{i \in \mathcal{I}} \left(Q(x^{[q]}, i) + \kappa(j, i) + W_{k+1}^{iT} \phi(f_i(x^{[q]})) \right), \forall q \in \{1, 2, \dots, p\}. \quad (21)$$

Step 7: Set $k = k - 1$. Go back to Step 5 until $k = 0$.

Algorithm 2 - Sequential Training

Step 1: For $j = 1$ to M repeat Step 2.

Step 2: Select an initial guess on W_N^j and repeat Steps 3 and 4 until W_N^j converges.

Step 3: Randomly select state vector $x \in \Omega$, where $\Omega \subset \mathbb{R}^n$ represents the domain of interest.

Step 4: Train weight W_N^j of neural network $W_N^j \phi(\cdot)$ using input-target pair $\{x, \psi(x, j)\}$.

Step 5: Set $k = N - 1$.

Step 6: For $j = 1$ to M repeat Step 7.

Step 7: Select an initial guess on W_k^j and repeat Steps 8 and 9 until W_k^j converges.

Step 8: Randomly select state vector $x \in \Omega$.

Step 9: Train W_k^j using input-target pair $\{x, \min_{i \in \mathcal{I}} (Q(x, i) + \kappa(j, i) + W_{k+1}^{iT} \phi(f_i(x)))\}$.

Step 10: Set $k = k - 1$. Go back to Step 6 until $k = 0$.

Remark 4: In order to have an idea of the computational load of the proposed method, one may consider the batch training form, Algorithm 1. The backward-in-time form of the algorithm resembles the solution to the conventional optimal control problem of discrete-time linear systems with quadratic cost functions, where, the Riccati (difference) equation needs to be evaluated for N time steps to find and store the time-varying solution, for all $k \in \mathcal{K}$, [24]. Eq. (20) resembles the final condition on the solution and Eq. (21) resembles the Riccati difference equation which takes the solution corresponding to $(k + 1)$ and provides the solution for time k . Due to the nonlinearity and the hybrid nature of the problem subject to this study, a function approximator needs to be utilized and several sample state vectors need to be selected at each evaluation of Eq. (21), for example, to find the unknown parameters. However, it can be seen that the computational load of the algorithm grows *linearly* as the number of time steps increases.

Finally, before concluding this section, it should be noted that the selection of *linear-in-parameter* form for the NN, as done in (8) and (9), is not required for the theory developed in this study to be valid. One can utilize *multi-layer perceptrons*, with k and i_{k-1} dependent weights, for improving the approximation capability of the NN. In this case, for example (9) changes to

$$NN(W_k^{i_{k-1}}, x_k) \approx V_k^*(x_k, i_{k-1}), \quad (22)$$

$$\forall k \in \mathcal{K} \cup \{N\}, \forall i_{k-1} \in \mathcal{I}, \forall x_k \in \Omega,$$

where function $NN(\cdot, \cdot)$ denotes the NN mapping, with the first argument being the tunable weights of the NN and the second argument being its input.

IV. ONLINE CONTROL

Once the NN is trained, it can be used for online control of the switching system. The process involves feeding the current state x_k and the already active subsystem i_{k-1} to equation

$$i_k^*(x_k, i_{k-1}) = \arg \min_{i \in \mathcal{I}} \left(Q(x_k, i) + \kappa(i_{k-1}, i) + W_{k+1}^{iT} \phi(f_i(x_k)) \right), \quad (23)$$

to find $i_k^*(x_k, i_{k-1})$. Note that, the minimization proposed in (23) is composed of comparing M scalar values and selecting the $i \in \mathcal{I}$ corresponding to the least value. Hence, the online *global* minimum can be easily found.

The advantages of this method are numerous. Firstly, the method provides a *feedback* solution, hence, it will be relatively robust toward uncertainties and disturbances. Secondly, no restriction is enforced on the order of the active subsystems or on the number of switching. Thirdly, unlike the nonlinear programming based methods [2]-[9] which give a local optimum, this method leads to an approximation of the global optimal solution. Note that this feature holds if the training of the NN, itself, is not stuck in a local minimum, which with the selection of linear-in-parameter NN and using convex methods like least squares, the condition is fulfilled. Fourthly, an important feature of this method is providing optimal switching for any initial condition $x_0 \in \Omega$ as long as the resulting state trajectory lies in the domain on which the network is trained, i.e., $x_k \in \Omega, \forall k$. The reason is the cost-to-go approximation is valid when the state belongs to Ω . Finally, the method provides a great deal of flexibility for implementation of different desired switching behaviors through admitting a general cost function with switching terms.

V. NUMERICAL ANALYSES

Two examples are selected for numerical investigation of the features of the proposed scheme. The source codes, in MATLAB, are available upon request.

A. Example 1

As the first example, a scalar problem with two modes, given below, is selected,

$$\dot{x} = f_1(x(t)) := -x(t), \quad \dot{x} = f_2(x(t)) := -x^3(t), \quad (24)$$

with the horizon of $2s$. For discretization of the continuous-time system, Euler forward integration, with a sampling time of $0.02s$ is selected which leads to $N = 100$. The selected cost function is

$$J = 5x_{100}^2 + \sum_{k=0}^{99} \kappa(i_{k-1}, i_k), \quad (25)$$

where

$$\kappa(i_{k-1}, i_k) = \begin{cases} 0 & \text{if } i_{k-1} = i_k \\ 0.1 & \text{if } i_{k-1} \neq i_k \end{cases} \quad (26)$$

Therefore, while the objective is bringing the state to close to zero, a cost of 0.1 is assumed for each switching. Hence, the controller should make a tradeoff between the cost due to the error in the state at the final time, and the cost due to switching. Considering the subsystems dynamics, both are stable. Comparing the derivatives of the state, however, subsystem 1 has a faster convergence rate when $|x| < 1$. But, when $|x| > 1$, subsystem 2 leads to a faster convergence of the state to the origin. Therefore, assuming there was no switching cost, the optimal solution would have been

$$i_k^* = \begin{cases} 1 & \text{if } |x| < 1 \\ 2 & \text{if } |x| > 1 \end{cases} \quad (27)$$

In comparing the neurocontroller results with Eq. (27), it should be noted that Eq. (27) is, loosely speaking, a “pointwise” optimal active mode, in the sense that it does not account for the cost of switching.

The basis functions were selected as polynomials x^j , where $j \in \{1, 2, \dots, 14\}$. The accuracy of the approximation capability of the NN can be adjusted by the selection of the order of the polynomials. The training was done over the domain of $\Omega = [-2, 2]$ in a sequential form and the weight were observed to converge in 1000 iterations. The resulting weight histories for the NN are plotted in Fig. 1. As expected, the weights are observed to be time-dependent, which represents the time-dependency of the cost-to-go.

After training the neurocontroller, initial condition $x_0 = 1.8$ was simulated using the developed method, for both cases of $i_{-1} = 2$ and $i_{-1} = 1$, i.e., the initial active mode being either subsystem 2 or subsystem 1. The results are given in Fig. 2. Considering the case of $i_{-1} = 2$, the utilized mode in the initial 18 time steps is *optimal* based on Eq. (27). Moreover, once the state becomes less than 1, if switching to mode 1 is eventually needed, i.e., the switching cost is unavoidable, then the switching should happen immediately based on Eq. (27), as done in Fig. 2.a. Comparing the cost-to-go 0.187 corresponding to Fig. 2.a with the cost-to-go of staying with mode 2 without any switching, which turned out to be 1.15, it is seen that the switching was required and the controller has provided optimal solution to the problem. Such an argument can be made for $i_{-1} = 1$ as well to analyze its optimality. The cost-to-go of the schedule given in Fig. 2.b turned out to be 0.197 which is less than the cost-to-go of operating mode 1 for the entire time (that is 0.297). Therefore, both switching, conducted in Fig. 2.b, were needed. Comparing the switching times with Eq. (27), the result given in Fig. 2.b is also optimal.

Note that the controller is able to provide optimal control for a vast number of initial conditions as long as the resulting state trajectory lies within Ω . From the dynamics of the subsystems it can be seen that selecting any $x_0 \in \Omega$ leads to $x_k \in \Omega, \forall k \in \mathcal{K}$. Therefore, the trained network can optimally control any initial condition $x_0 \in \Omega$. The initial condition $x_0 = 1.3$ was selected next. The results are presented in Fig. 3. While Fig. 3.a and comparing its cost-to-go, 0.146, with the cost-to-go of no switching at all, 1.084, show that the controller has optimally switched between the modes per Eq. (27), an interesting observation can be made from Fig. 3.b. As seen in this figure, once the initial active mode is

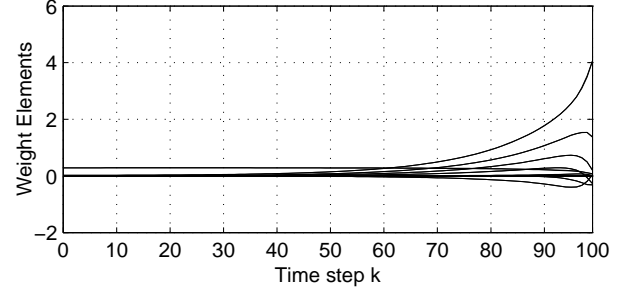


Fig. 1. History of elements of the weight of the trained NN.

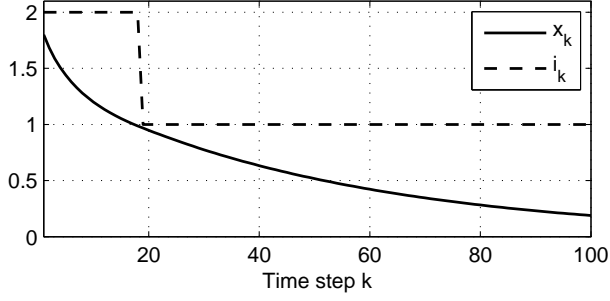
not “pointwise” optimal, but, the initial condition is not large enough such that switching to the pointwise optimal mode leads to an enough *reward* to cover the cost of switching, the controller stays with the initial mode and skips the switching. This can be confirmed by comparing the cost-to-go of the case of switching from mode 1 to mode 2 at the very beginning and switching back to mode 1 right when x becomes smaller than 1, which turned out to be 0.156, with the cost of staying with mode 1 for the entire time, that is 0.155. Since the cost-to-go of the former case is more than that of the latter case, no switching was needed and the controller has optimally controlled the new initial condition.

A very important feature of the solution is the fact that this method does not *postpone* switching instants, as the remedy proposed in [28] does. If a switching is eventually needed, then it should happen at the *best* time without spending time with operating the non-optimal mode. For example, in Figs. 2.a and 3.a, where the initial active mode is 2 and a switching is eventually needed, the switching has happened right at the best time, i.e., the time that the state became less than 1.

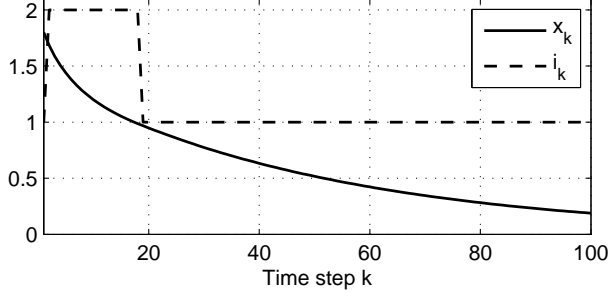
Finally, the initial condition $x_0 = 0.8$ is simulated for both the two initial active modes, and the results are depicted in Fig. 4. Considering the history of the active modes, the neurocontroller has optimally controlled the system through either not switching at all, or switching immediately, considering the state histories.

B. Example 2

A nonlinear second order system with three modes, simulated in [7] and [28], is selected as the second example. The objective of this problem is controlling the fluid level in a two-tank setup. The fluid flow into the upper tank can be adjusted through a valve which has three positions: fully open, half open, and fully closed. Each tank leaks fluid with a rate proportional to the square root of the height of the fluid in the respective tank. The upper tank leaks into the lower tank, and the lower tank leaks to the outside of the setup. Representing the fluid height in the upper tank with scalar y and in the lower tank with scalar z , the dynamics of the state vector $x = [y, z]^T$ are given by the following three modes, corresponding to the

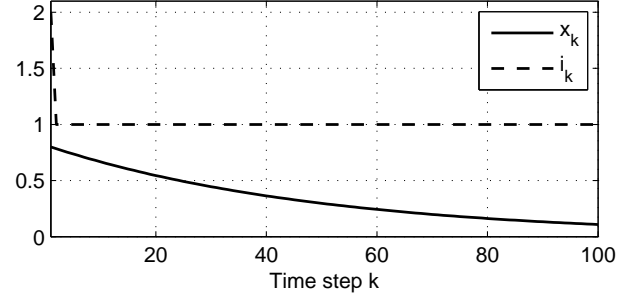


a: Initial mode $i_{-1} = 2$.

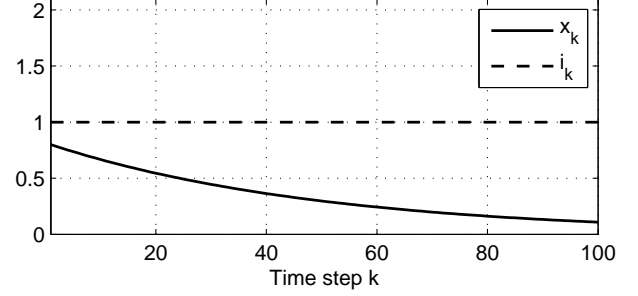


b: Initial mode $i_{-1} = 1$.

Fig. 2. History of state and active mode for initial condition $x_0 = 1.8$.

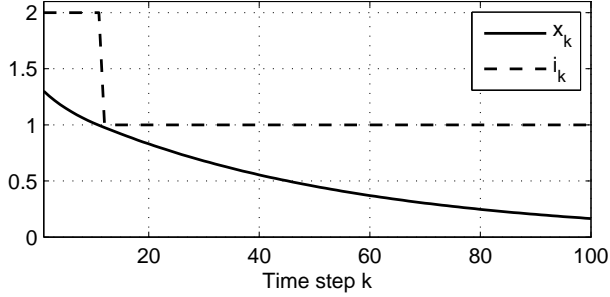


a: Initial mode $i_{-1} = 2$.

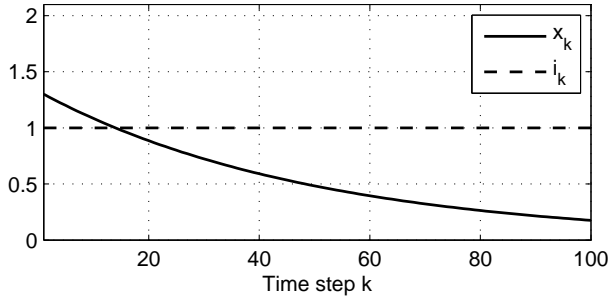


b: Initial mode $i_{-1} = 1$.

Fig. 4. History of state and active mode for initial condition $x_0 = 0.8$.



a: Initial mode $i_{-1} = 2$.



b: Initial mode $i_{-1} = 1$.

Fig. 3. History of state and active mode for initial condition $x_0 = 1.3$.

three positions of the valve,

$$\begin{aligned} \dot{x} &= f_1(x) := \begin{bmatrix} -\sqrt{y} \\ \sqrt{y} - \sqrt{z} \end{bmatrix}, \\ \dot{x} &= f_2(x) := \begin{bmatrix} -\sqrt{y} + 0.5 \\ \sqrt{y} - \sqrt{z} \end{bmatrix}, \\ \dot{x} &= f_3(x) := \begin{bmatrix} -\sqrt{y} + 1 \\ \sqrt{y} - \sqrt{z} \end{bmatrix}. \end{aligned} \quad (28)$$

The objective is forcing the fluid level in the lower tank, i.e., z , to track constant value 0.5. Selecting the control horizon $5s$ the problem was discretized using sampling time of $0.025s$, therefore $N = 200$. Then, cost function (2) was selected for evaluating the performance of the method in both decreasing the number of switching and also for assigning certain preferences in utilization of some modes. The basis functions for this example were selected as polynomials $y^p z^q$, where non-negative integers p and q are such that $p + q \leq 8$. This selection led to 45 neurons. Domain $\Omega = \{[y, z]^T \in \mathbb{R}^2 : 0 \leq y < 1, 0 \leq z < 0.8\}$ was used for the training and batch training scheme was selected, such that, at each stage 100 random states were selected based on Algorithm 1.

Initially the following values were selected for the cost function

$$\begin{aligned} \psi(x_N, i_{N-1}) &= 0.25(z_N - 0.5)^2, \forall i_{N-1} \in \mathcal{I}, \\ Q(x_k, i_k) &= 0.25(z_k - 0.5)^2, \forall i_k \in \mathcal{I}, \\ \kappa(i_{k-1}, i_k) &= \begin{cases} 0 & \text{if } i_{k-1} = i_k \\ \kappa_0 & \text{if } i_{k-1} \neq i_k \end{cases}, \end{aligned} \quad (29)$$

with $\kappa_0 = 0$. As seen, such a cost function does not assign any cost to switching and does not differentiate between the modes. The training was observed to take almost 16 seconds, when $N = 200$, in a machine with CPU Intel Core i7, 3.4 GHz running MATLAB 2013a. Afterwards, initial condition $x_0 = [0.8, 0.2]^T$, simulated in [7] and [28], was used to determine the optimal solution. Selecting $i_{-1} = 3$, the results are given in Fig. 5. The method did an excellent job controlling the fluid level of the lower tank by tracking the desired value. This perfect tracking was achievable, however, through high frequency switching between the three modes. Next, the switching cost $\kappa_0 = 0.001$ was used with terms given in (29).

The training was re-done using the new cost function and the simulation results in controlling the same initial condition are shown in Fig. 6. It is seen that the incorporated switching cost has effectively lowered the number of switching, compared with Fig. 5. Assigning higher cost to switching, like $\kappa_0 = 0.01$, further decreases the number of switching while tracking 0.5, as shown in Fig. 7.

The application of switching costs for decreasing the number of switching resembles the idea utilized in [28] and called *Threshold Remedy*, in which no switching cost was incorporated, and hence, the optimal cost-to-go was approximated as a standard function of only the state and the time, as in non-switching problems [21]. The training also was done without including a switching cost. In online control process, however, a threshold, similar to the switching cost used in this study, was applied. This was done in the sense that, the reward of switching to another mode must be higher than a certain threshold in order for the controller to switch. While this remedy can help in certain conditions, the performance deviates as the threshold becomes large. The reason is, the threshold is not accounted for in the learning process and hence, the result given in [28] is not optimal considering the applied threshold. To see this, one may compare Fig. 7 with the results given in Fig. 8, where the former is the result of the proposed method in this study with $\kappa_0 = 0.01$ and the latter is the result obtained from [28] with the equal threshold of 0.01, in lieu of the switching cost. As seen, the tracking is much more accurate in Fig. 7 versus Fig. 8. Even considering the cost due to the extra switches in Fig. 7, the cost-to-go corresponding to Fig. 7 turned out to be 0.340 which is less than the cost-to-go corresponding to Fig. 8, i.e., 0.468. High threshold values can potentially lead to unreliability of the result of the method in [28]. This problem does not exist with the method presented here due to the fact the switching cost is incorporated in the derivation and the solution is optimal with respect to it.

Once the performance of the controller in applying switching costs is analyzed, the cost function terms $\psi(\cdot, \cdot)$ and $Q(\cdot, \cdot)$ are modified to assign certain *mode preferences* in the operation of the system. The mode preference is using modes 2 and 3 more often *during* the operation and *finishing* the operation preferably with mode 1. This was done through selecting the following cost function terms

$$\begin{aligned} \psi(x_N, i_{N-1}) &= \begin{cases} 0.25(z_N - 0.5)^2 - 10, & \text{if } i_{N-1} = 1 \\ 0.25(z_N - 0.5)^2, & \text{if } i_{N-1} \neq 1 \end{cases}, \\ Q(x_k, i_k) &= \begin{cases} 0.25(z_k - 0.5)^2 + 0.01, & \text{if } i_k = 1 \\ 0.25(z_k - 0.5)^2, & \text{if } i_k \neq 1 \end{cases} \end{aligned} \quad (30)$$

along with the switching cost $\kappa(i_{k-1}, i_k)$ given in (29) with $\kappa_0 = 0$. Note that the negative cost -10 assigned to $\psi(\cdot, 1)$ provides the controller with rewards if the last active mode is 1. Also, the positive cost 0.01 assigned to $Q(\cdot, 1)$ penalizes the usage of mode 1 during the horizon. Moreover, it should be noted that the method presented in this study does not require the cost function terms to be positive semi-definite, see Remark 1. Hence, one can select smooth functions with negative values as well as positive semi-definite functions.

Having trained the neurocontroller based on the new cost function terms, the results for controlling the same initial x_0 , with $i_{-1} = 1$, are presented in Fig. 9. Comparing this figure with Fig. 5, it is seen that the new controller has effectively decreased the number of instants that mode 1 is used. Also, looking at the final active mode in Fig. 9, the controller has been successful in finishing the operation with mode 1, as desired. As another simulation, the cost function terms given in (30) with the switching cost $\kappa_0 = 0.0002$ is used for training the network and the simulation results are given in Fig. 10. This figure demonstrates the capability of the method in both lowering the number of switching and assigning different preferences to utilization of different modes.

Next, the capability of the neurocontroller in controlling different initial conditions within Ω is investigated. A new initial condition, namely $x_0 = [0, 0]^T$ is utilized as the last simulation and the last trained network (without re-training) is used for controlling it. The results, given in Fig. 11, show the capability of the controller in controlling the new initial condition with the desired manner without any need for retraining. In order to furthermore evaluate this capability, several different initial conditions are selected and the resulting histories for z_k , which is the variable of interest, from simulating each of the initial conditions are presented in Fig. 12. In these plots, initially z_0 is fixed at 0.2 and y_0 is changed from 0 to 1 in steps of 0.1, and then, y_0 is fixed at 0.8 and z_0 is changed from 0 to 0.8, in steps of 0.1. The plots show that the controller has effectively controlled all these different initial conditions through tracking the desired constant value 0.5, without re-training.

Finally, the closed-loop feature of the proposed method is evaluated in terms of its moderate robustness. A time-varying random disturbance is introduced after the training phase in the online operation and is assumed to be uniformly changing between 0 and 0.005, acting as additive terms on both state elements. The resulting history for the variable of interest z_k , is depicted in Fig. 13 through the solid plot. Moreover, the history if the system was operated using an open loop solution is also included, through the dash plot in the same figure. As seen, the system operated in the open loop fashion, that is, when the mode sequence in Fig. 10, which was calculated under no disturbance situation is applied, leads to instability of the system, but, the closed from solution handles the disturbance with a slight performance degradation in terms of a small steady state error. This demonstrates the desired feature of the proposed method.

VI. CONCLUSIONS

The problem of finding the optimal switching schedule between different modes of a dynamical system with a cost function which admits incorporation of switching costs as well as assigning different costs to different modes was investigated and the framework of approximate dynamics programming was used with the idea of approximating the optimal cost-to-go for solving it. It was shown that for such problems the cost-to-go function is not only a function of the current state and time, but also, a function of the subsystem which was active at the previous time step. It was shown that the developed technique

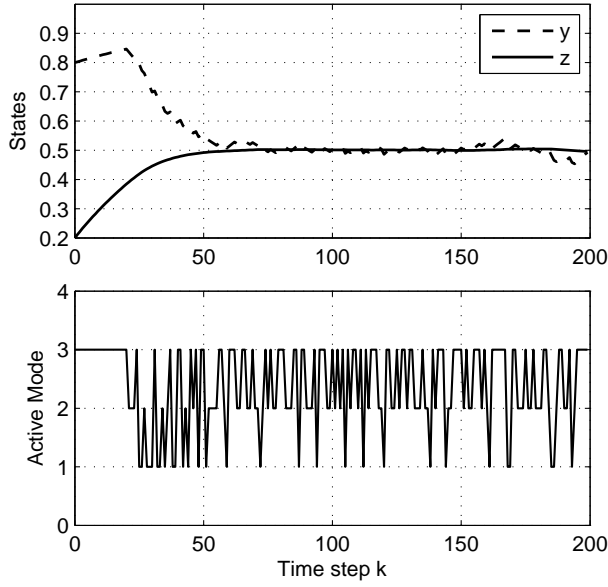


Fig. 5. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0$, and cost function terms given in Eq. (29).

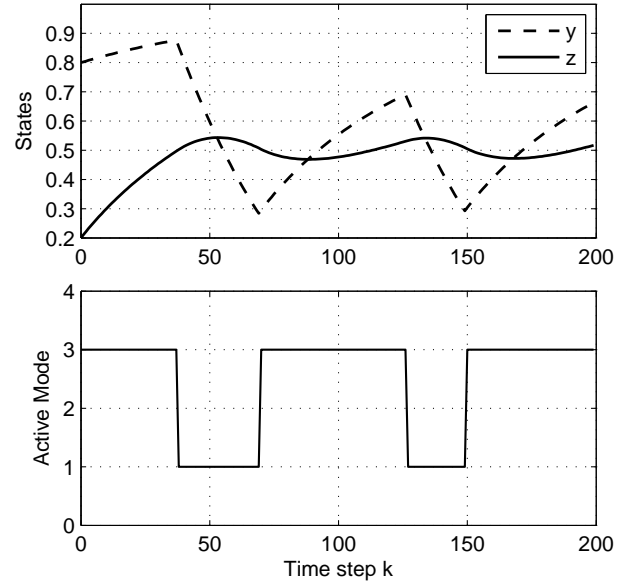


Fig. 7. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0.01$, and cost function terms given in Eq. (29).

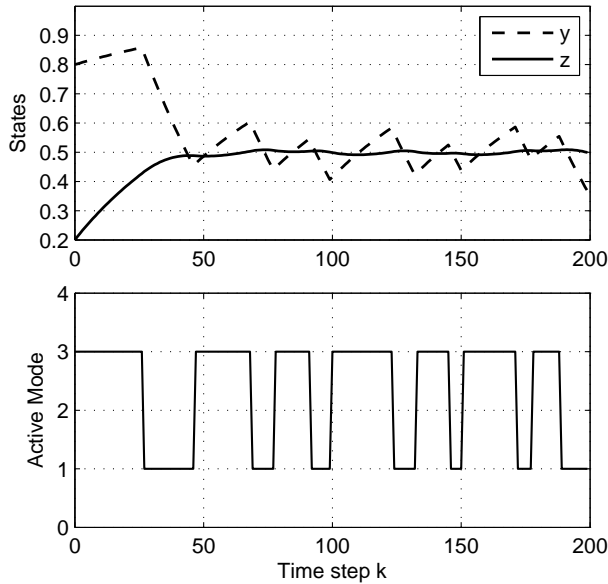


Fig. 6. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0.001$, and cost function terms given in Eq. (29).

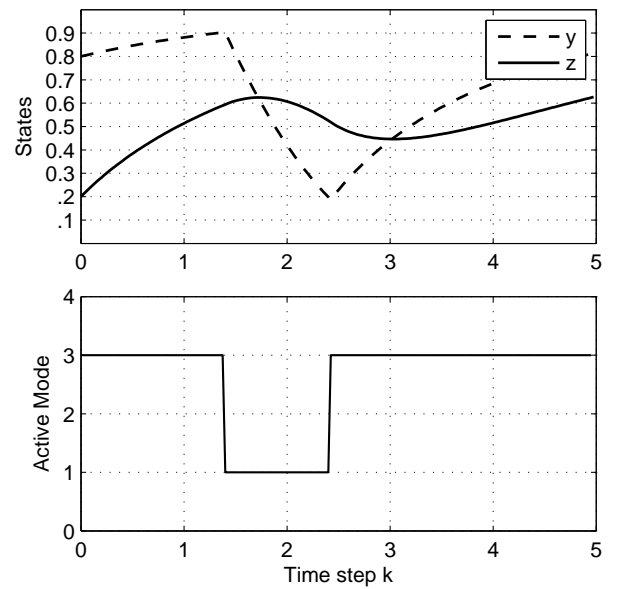


Fig. 8. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0$, and cost function terms given in Eq. (29) with applied threshold 0.01 proposed in [28].

can effectively provide (approximate) optimal solutions to the problems with different initial conditions in a feedback form. The real-time computational burden of the method is as small as evaluating as many scalar-valued functions as the number of subsystems.

REFERENCES

- [1] A. Heydari, and S. N. Balakrishnan, "Optimal orbit transfer with on-off actuators using a closed form optimal switching scheme," *AIAA Guidance, Navigation, and Control Conference*, Boston, MA, 2013.
- [2] X. Xu, and P. J. Antsaklis, "Optimal control of switched systems via nonlinear optimization based on direct differentiations of value functions," *International Journal of Control*, vol. 75 (16/17), pp. 1406-1426, 2002.
- [3] X. Xu, and P. J. Antsaklis, "Optimal control of switched systems based on parameterization of the switching instants," *IEEE Trans. on Automatic Control*, vol. 49 (1), pp.2- 16, 2004.
- [4] M. Egerstedt, Y. Wardi, and H. Axelsson, "Transition-time optimization for switched-mode dynamical systems," *IEEE Trans. on Automatic Control*, vol. 51 (1), pp.110-115, 2006.
- [5] H. Axelsson, M. Boccadoro, M. Egerstedt, P. Valigi, and Y. Wardi, "Optimal mode-switching for hybrid systems with varying initial states," *Nonlinear Analysis: Hybrid Systems*, vol. 2 (3), pp.765772, 2008.
- [6] X. Ding, A. Schild, M. Egerstedt, J. and Lunze, "Real-time optimal feedback control of switched autonomous systems," *Proc. IFAC Conference on Analysis and Design of Hybrid Systems*, pp.108-113, 2009.
- [7] Axelsson, H., Egerstedt, M., Wardi, Y., and Vachtsevanos, G., "Algorithm for switching-time optimization in hybrid dynamical systems," *Proc. IEEE International Symposium on Intelligent Control*, Limassol, Cyprus, 2005.

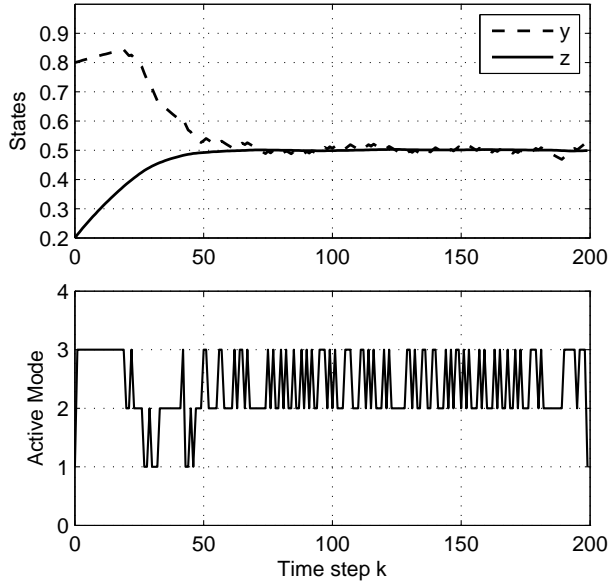


Fig. 9. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0$, and cost function terms given in Eq. (30).

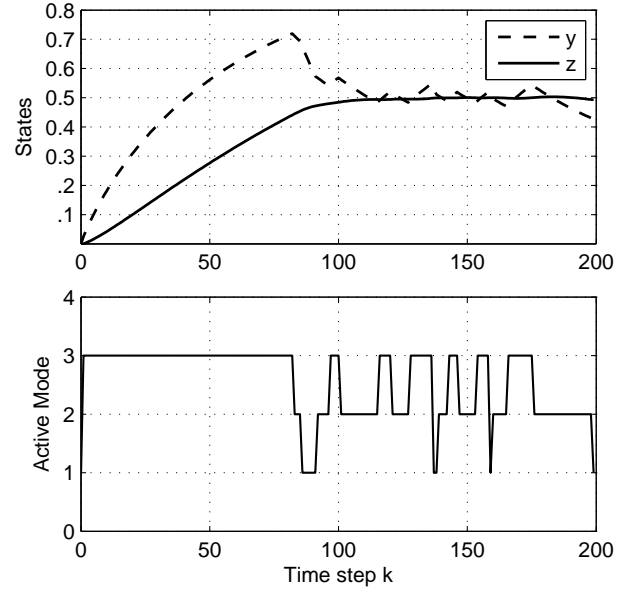


Fig. 11. Simulation result for $x_0 = [0, 0]^T$, $\kappa_0 = 0.0002$, and cost function terms given in Eq. (30).

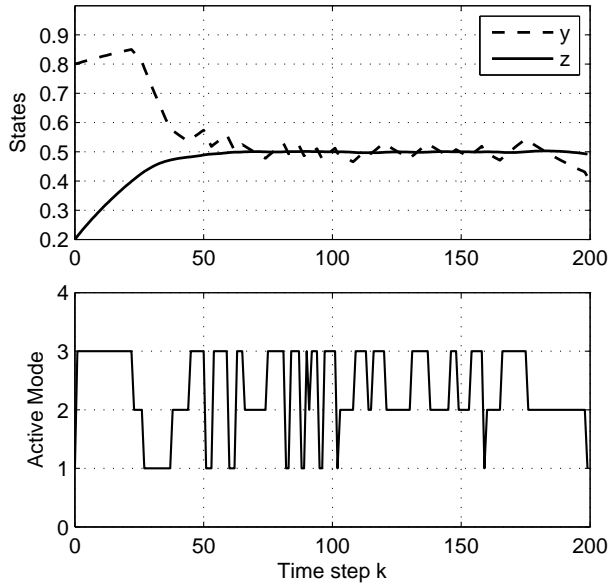


Fig. 10. Simulation result for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0.0002$, and cost function terms given in Eq. (30).

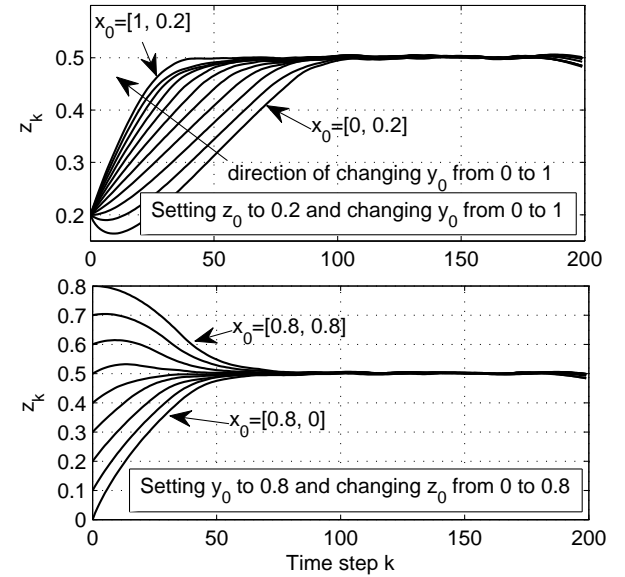


Fig. 12. Simulation result for different initial conditions with $\kappa_0 = 0.0002$, and cost function terms given in Eq. (30).

- [8] M. Kamgarpoura, and C. Tomlin, "On optimal control of non-autonomous switched systems with a fixed mode sequence," *Automatica*, vol. 48, pp.1177-1181, 2012.
- [9] R. Zhao, and S. Li, "Switched system optimal control based on parameterizations of the control vectors and switching instant," *Proc. Chinese Control and Decision Conference*, pp. 3290-3294, 2011.
- [10] M. Rungger, and O. Stursberg, "A numerical method for hybrid optimal control based on dynamic programming," *Nonlinear Analysis: Hybrid Systems*, vol. 5 (2), pp.254-274, 2011.
- [11] H. Zhang, and M. R. James, "On computation of optimal switching HJB equation," *IEEE Conference on Decision and Control*, 2006, pp. 2704-2709.
- [12] M. Sakly, A. Sakly, M. Majdoub, and M. Benrejeb, "Optimization of switching instants for optimal control of linear switched systems based on genetic algorithms," *Proc. IFAC Int. Conf. Intelligent Control Systems and Signal Processing*, Istanbul, 2009.

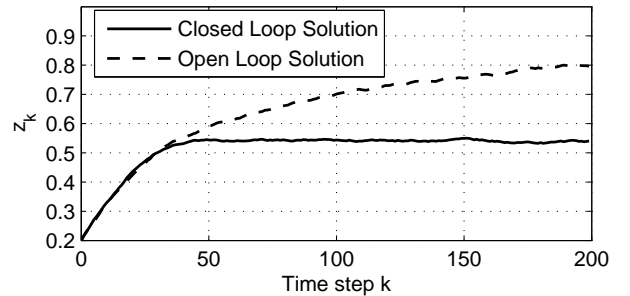


Fig. 13. Simulation result under random time-varying disturbance for $x_0 = [0.8, 0.2]^T$, $\kappa_0 = 0.0002$, and cost function terms given in Eq. (30).

- [13] R. Long, J. Fu, L. Zhang, "Optimal control of switched system based on neural network optimization," *Proc. Int. Conference on Intelligent Computing*, pp.799-806, 2008.
- [14] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling". In D. A. White, and D. A. Sofge (Eds.), *Handbook of Intelligent Control*, Multiscience Press, 1992.
- [15] S. N. Balakrishnan, and V. Biega, "Adaptive-critic based neural networks for aircraft optimal control", *Journal of Guidance, Control and Dynamics*, vol. 19 (4), 1996, pp. 893-898.
- [16] D. V. Prokhorov, and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Networks*, vol. 8 (5), 1997, pp. 997-1007.
- [17] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof," *IEEE Trans. Systems, Man, and Cybernetics-Part B*, vol. 38, 2008, pp. 943-949.
- [18] G. K. Venayagamoorthy, R. G. Harley, and D. C. Wunsch, "Comparison of heuristic dynamic programming and dual heuristic programming adaptive critics for neurocontrol of a turbogenerator," *IEEE Trans. Neural Netw.*, vol. 13 (3), pp. 764-773, 2002.
- [19] T. Dierks, B. T. Thumati, and S. Jagannathan, "Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence," *Neural Networks*, vol. 22, pp. 851-860, 2009.
- [20] Vrabie, D. and Vamvoudakis, K.G. and Lewis, F.L., *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*, IET control engineering series, Institution of Engineering and Technology, 2013.
- [21] A. Heydari, S. N. Balakrishnan, "Fixed-final-time optimal control of nonlinear systems with terminal constraints," *Neural Network*, vol. 48, pp. 61-71, 2013.
- [22] A. Heydari, S. N. Balakrishnan, "Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24 (1), pp. 145-157, 2013.
- [23] F. Wang, N. Jin, D. Liu, and Q. Wei, "Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ϵ -error bound," *IEEE Trans. Neural Netw.*, vol. 22 (1), pp. 24-36, 2011.
- [24] D. E. Kirk, *Optimal Control Theory: An Introduction*, Dover Publications, New York, 2004, pp. 53-58.
- [25] A. Heydari, S. N. Balakrishnan, "Optimal switching and control of nonlinear switched systems using approximate dynamic programming," to appear in *IEEE Transactions on Neural Networks and Learning Systems*, 2014.
- [26] A. Heydari, S. N. Balakrishnan, "Optimal multi-therapeutic HIV treatment using a global optimal switching scheme," *Applied Mathematics and Computation*, vol. 219, pp. 7872-7881, 2013.
- [27] A. Heydari, S. N. Balakrishnan, "Optimal switching between controlled subsystems with free mode sequence," submitted to *Neurocomputing*.
- [28] A. Heydari, S. N. Balakrishnan, "Optimal switching between autonomous subsystems," *Journal of the Franklin Institute*, Vol. 351, pp. 2675-2690, 2014.
- [29] C. Qin, H. Zhang, Y. Luo, and B. Wang, "Finite horizon optimal control of non-linear discrete-time switched systems using adaptive dynamic programming with ϵ -error bound," *International Journal of Systems Science*, 2013.
- [30] W. Lu, S. Ferrari, "An approximate dynamic programming approach for model-free control of switched systems," *Proceedings of the IEEE Conference on Decision and Control*, pp. 3837-3844, 2013.
- [31] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.
- [32] R. Courant, and D. Hilbert, "Methods of Mathematical Physics," Vol I. Wiley (Interscience), New York, p. 65, 1953.
- [33] W. F. Trench, *Introduction to Real Analysis*, available online at http://ramanujan.math.trinity.edu/wtrench/texts/TRENCH_REAL_ANALYSIS.PDF, 2012, p. 309.